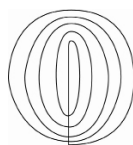


CONSEQUENCIALISMO DAS REGRAS

EDIÇÃO DE 2019 do

COMPÊNDIO EM LINHA  
DE PROBLEMAS DE FILOSOFIA ANALÍTICA

2018-2021 FCT Project PTDC/FER-FIL/28442/2017



Editado por  
Ricardo Santos e Pedro Galvão

ISBN: 978-989-8553-22-5

Compêndio em Linha de Problemas de Filosofia Analítica  
Copyright © 2019 do editor  
Centro de Filosofia da Universidade de Lisboa  
Alameda da Universidade, Campo Grande, 1600-214 Lisboa

Consequencialismo das Regras  
Copyright © 2019 do autor  
Pedro Galvão

Todos os direitos reservados

# Consequencialismo das regras

**Resumo:**

Segundo o consequencialismo das regras, devemos seguir o código moral que produziria as consequências imparcialmente melhores. Neste artigo, distinguem-se e discutem-se diversas versões de consequencialismo das regras. De seguida, examinam-se dois argumentos importantes a favor da teoria: o Argumento da Equiprobabilidade, de Harsanyi, e o Argumento Kantiano, de Parfit. Finalmente, avalia-se o consequencialismo das regras em função do seu acordo com intuições morais e da sua capacidade de explicar o estatuto deôntico dos actos.

**Palavras-chave:**

consequencialismo, ética kantiana, contratualismo, deontologia, códigos morais

**Abstract:**

According to rule consequentialism, we ought to follow the moral code that would produce the impartially best consequences. In this article, several versions of rule consequentialism are distinguished and discussed. Then two important arguments for the theory are examined: Harsanyi's Equiprobability Argument, and Parfit's Kantian Argument. Finally, rule consequentialism is assessed according to its agreement with moral intuitions and its ability to explain the deontic status of acts.

**Keywords:**

Consequentialism, Kantian ethics, contractualism, deontology, moral codes

# Consequencialismo das Regras

## 1 Introdução

Devemos seguir os princípios (ou regras) que resultem nas consequências imparcialmente melhores. É isto que, no essencial, nos diz o consequencialismo das regras. Sendo uma teoria da obrigação, ou do dever, esta perspectiva oferece-nos um padrão de *correção moral*, isto é, um critério para determinar que actos são correctos (ou moralmente aceitáveis) e que actos são incorrectos (ou moralmente errados). Ao mesmo tempo, visa explicar-nos o que torna correctos os actos que são correctos, e incorrectos os actos que são incorrectos. Por outras palavras, visa explicar-nos o *estatuto deontico* de actos particulares.

De acordo com uma classificação comum, o consequencialismo das regras inclui-se entre as formas *indirectas* de consequencialismo. Estas formas contrastam com o consequencialismo dos actos, que nos diz para avaliar directamente cada acto particular em termos da promoção do bem. Na sua versão mais difundida, esta perspectiva afirma que um acto moralmente aceitável é aquele que terá consequências imparcialmente melhores, pelo que um acto será errado se existir um acto alternativo cuja realização resulte num maior bem.

O consequencialista das regras discorda. Por considerar que o consequencialismo dos actos tem implicações práticas demasiado implausíveis, entende ser um erro usar o padrão da promoção imparcial do bem para avaliar directamente actos particulares. Essa avaliação directa deverá incidir não em actos particulares alternativos, mas em princípios alternativos — ou, melhor ainda, em conjuntos de princípios, isto é, em códigos morais alternativos. O *código moral* correcto, ou *ideal*, será aquele que, sendo constituído pelos *princípios optimíficos*, resultará em melhores consequências do que qualquer outro código moral. O estatuto deontico dos actos particulares será, por sua vez, uma questão de concordância com o código ideal. Este é, pois, o padrão que define o consequencialismo das regras:

**CR.** Um acto é moralmente errado se, e apenas se, infringe os princípios optimíficos.

Aquilo que torna um acto errado ou incorrecto, importa acrescentar,

Publicado pela primeira vez em 2019

é o facto de este infringir os princípios do código ideal. Um acto que seja errado será errado *sempre* em virtude de ter a propriedade de infringir os princípios optimíficos — o que, note-se desde já, não implica que seja errado *apenas* em virtude de ter essa propriedade.

O padrão do consequencialista das regras será vazio na ausência de uma concepção do bem, mais precisamente de uma concepção daquilo que é intrinsecamente melhor ou pior acontecer, *vistas as coisas de uma perspectiva imparcial*. Só à luz de uma concepção suficientemente definida do bem será possível avaliar as consequências de códigos morais alternativos. Como seria de esperar, não há uma concepção do bem que seja consensual entre os consequencialistas das regras. Alguns, como John C. Harsanyi (1977, 1985) e Richard Brandt (1979), são *utilitaristas*, o que significa que concebem o bem simplesmente como o bem-estar agregado. Para outros, como Brad Hooker (2000) e Derek Parfit (2011a), o bem depende também de factores aos quais o utilitarista não atribui importância intrínseca, como uma determinada *distribuição* do bem-estar. A este respeito, Hooker tem uma posição mais definida que Parfit, pois defende que os princípios optimíficos são aqueles que melhor promovem dois valores distintos: (1) o bem-estar agregado *total* (e não médio), e (2) a equidade [*fairness*] na distribuição de bem-estar, que será maior quanto melhor for a situação dos mais desfavorecidos.

Os quatro filósofos mencionados perfazem a lista dos consequencialistas das regras mais influentes. Não é uma lista longa. Na verdade, não se pode dizer que o consequencialismo das regras alguma vez tenha sido objecto de uma ampla simpatia. Para muitos, esta teoria representa um compromisso insustentável entre, por um lado, o consequencialismo comum, dos actos, e, por outro, as éticas kantianas ou contratualistas, que também concebem o estatuto deontico dos actos como uma questão de concordância com um determinado código ideal — ainda que, para a generalidade dos kantianos e dos contratualistas, esse código *não* seja constituído pelos princípios optimíficos. Que é injustificada (ou, no mínimo, precipitada) a alegação de que o consequencialismo das regras exemplifica um compromisso *insustentável*, tornar-se-á claro nas próximas secções.

## 2 Cumprimento e aceitação de princípios

Por si mesmos, princípios ou regras morais não têm consequências. Só têm consequências, melhores ou piores, quando relacionados apropriadamente com agentes. Para avaliar princípios à luz do padrão consequencialista, que relação, então, haveremos de supor que se verifica entre estes e os agentes morais? Às duas respostas principais a esta questão correspondem duas versões de consequencialismo das regras, que diferem na concepção dos princípios optimíficos. Segundo uma delas, os princípios optimíficos são aqueles que resultam nas melhores consequências quando colectivamente *cumpridos* (ou observados) pelos agentes. De acordo com a outra, esses princípios são antes aqueles que resultam nas melhores consequências quando colectivamente *aceites* (ou interiorizados) pelos agentes.

Para esclarecer a diferença entre cumprir e aceitar um princípio, consideremos, a título de exemplo, um princípio extremamente simples: mentir é errado. Cumprir esta regra é apenas uma questão de não ter uma conduta mentirosa. Aceitá-la ou tê-la interiorizado, pelo contrário, consiste em ter determinadas disposições, subsumíveis numa aversão geral à mentira. Esta aversão inclui, obviamente, uma disposição para não mentir, mas também disposições para censurar a conduta mentirosa, para evitar mentirosos, para aprovar uma educação moral que repudie a mentira ou para sentir alguma forma de remorso caso se minta. Como este último aspecto deixa claro, é possível aceitar um princípio mas nem sempre o cumprir. Inversamente, também é possível cumprir um princípio sem o aceitar, pois é concebível que um agente acabe por nunca mentir, apesar de não ter nenhuma aversão à mentira.

Visto que os princípios têm consequências relevantes para lá das que decorrem do seu simples cumprimento, o consequencialismo das regras, na sua versão de referência, afirma que devemos avaliar princípios pensando nas consequências da sua *aceitação*. Sejam  $C_1$  e  $C_2$  códigos morais alternativos. Há que considerar, então, o cenário no qual cada um destes códigos é colectivamente aceite. *Ceteris paribus*, o código moral preferível será aquele que, nesse tipo de situação possível, teria melhores consequências. *Ceteris paribus*, se  $C_1$  teria melhores consequências do que  $C_2$ , então será preferível a  $C_2$ . Se  $C_1$  teria melhores consequências do que qualquer outro

código alternativo, então será o código ideal. A cláusula *ceteris paribus* tem uma fortíssima razão de ser. Qualquer código moral tem custos de interiorização. Ou seja, levar cada geração de seres humanos a interiorizar um dado conjunto de princípios morais tem custos tanto cognitivos como motivacionais. Estes custos, muito variáveis, serão tanto mais elevados quanto mais complexo ou mais exigente for um código moral. A avaliação de códigos morais não deverá ser-lhes insensível: *ceteris paribus*, os códigos morais com menores custos de interiorização são preferíveis. Ora, suponha-se que, após a sua interiorização, C<sub>1</sub> teria consequências um pouco melhores do que C<sub>2</sub>, mas que C<sub>1</sub> tem custos de interiorização muitíssimo superiores aos de C<sub>2</sub>. Nestas circunstâncias, ponderadas todas as coisas, o melhor código moral não seria C<sub>1</sub>, mas C<sub>2</sub>.

Dadas as considerações precedentes, podemos avançar de CR para uma versão mais definida de consequencialismo das regras:

**CR\*.** Um acto é moralmente errado se, e apenas se, infringe os princípios cuja aceitação colectiva resultaria em melhores consequências, descontados os seus custos de interiorização.

Há que pressupor a possibilidade psicológica de esses princípios serem aceites por agentes humanos. Não havendo essa possibilidade, os seus custos de interiorização seriam como que infinitos. Examinemos agora uma das objecções mais salientes ao consequencialismo das regras — a de que este, na verdade, ‘colapsa’ no consequencialismo dos actos, pelo que não é uma alternativa genuína a esta teoria.<sup>1</sup> O alegado ‘colapso’ consiste na ideia de que, de uma forma ou de outra, o código ideal, segundo o padrão do consequencialista das regras, será extensionalmente equivalente ao padrão de correcção que o consequencialista dos actos subscreve. Por outras palavras, ao invés do que os consequencialistas das regras presumem, esse código dir-nos-á que os actos errados são precisamente aqueles que não têm as melhores consequências, por comparação com os actos alternativos.

De acordo com esta objecção, o código ideal tem de consistir simplesmente num princípio como ‘Devemos fazer sempre aquilo que resulte no maior bem’ — ou então num conjunto mais complexo

<sup>1</sup> Veja-se Smart (1973: 11-12). Para uma réplica à objecção nas suas diversas variantes, veja-se Hooker (2000: 93-99).

de princípios, que, no entanto, equivalerão à exigência de maximizar o bem em todas as circunstâncias. Por que razão haveremos de julgar que o código ideal tem este conteúdo? Porque, se todos o seguirem, o resultado será a maximização imparcial do bem. Não há, pois, um código moral com consequências ainda melhores.

Talvez esta objecção seja bem-sucedida caso se dirija a uma versão de consequencialismo das regras que respeite ao *cumprimento* de princípios. Se o código constituído pelo princípio ‘Devemos fazer sempre aquilo que resulte no maior bem’ — chamemos-lhe CA — for colectivamente cumprido, cada um fará sempre aquilo que terá as melhores consequências e, dessa forma, o bem será imparcialmente maximizado, pelo que não existirá um código moral com consequências ainda melhores.

Ainda que isto possa ser verdade, a objecção falha se o seu alvo for CR\*, pois há razões decisivas para crer que o código moral cuja *aceitação* colectiva teria melhores consequências *não é* CA. Antes de mais, consideremos os custos de interiorização deste código moral. Embora CA seja muitíssimo simples, tendo por isso custos cognitivos insignificantes, os seus custos motivacionais são inimaginavelmente altos. Na verdade, o mais certo é a sua aceitação colectiva por seres humanos ser nada menos que psicologicamente *impossível*. É que ter interiorizado o código moral constituído pela regra ‘Devemos fazer sempre aquilo que resulte no maior bem’ corresponde a ter uma só motivação: agir por beneficência universal, procurando promover o bem com a imparcialidade mais estrita em todas as circunstâncias. O extremo altruísmo necessário ao desenvolvimento desta motivação pura e simplesmente não se coaduna com a natureza humana.

Conceda-se, no entanto, a possibilidade psicológica da aceitação colectiva de CA em cada geração de seres humanos. Contrariamente ao que poderá parecer óbvio, CA, nesse cenário de aceitação, não teria as melhores consequências. Por duas razões distintas, na verdade teria consequências desastrosas. Em primeiro lugar, repare-se que o princípio de CA, embora seja simples no seu conteúdo, é extremamente difícil de *aplicar*. Determinar, em cada caso, que acto terá as melhores consequências requer capacidades de previsão que excedem enormemente as humanas. Assim, se CA conquistasse a aceitação colectiva, o princípio ‘Devemos fazer sempre aquilo que resulte no maior bem’ normalmente seria mal aplicado, pelo que o

bem não seria maximizado. Em segundo lugar, importa reconhecer que muito do que há de valioso, contribuindo inestimavelmente para o bem-estar dos indivíduos, depende de parcialidades. Isto é inegável no que concerne às relações pessoais profundas, de amizade ou de amor, visto que estas só podem existir entre os seres humanos caso estes se importem muito especialmente com alguns indivíduos. Assim, numa situação em que os agentes estivessem motivados exclusivamente pela beneficência universal, essas relações pessoais não seriam possíveis. Além do mais, a dedicação a projectos espontaneamente pessoais (artísticos, lúdicos, científicos, económicos, etc.) também ficaria comprometida, resultando daí outra perda imensa de valor. A aceitação colectiva de uma exigência ilimitada de maximizar o bem, em suma, levaria seguramente a consequências bem piores do que as de alguns códigos alternativos mais lenientes em matéria de beneficência universal.

### 3 Identificação dos princípios optimíficos

Se o consequencialismo das regras não conduz a CA, qual será então o código ideal? Quais são, na verdade, os princípios optimíficos? Poderá parecer que, para identificarmos o código ideal, teremos de proceder deste modo:

- (1) Formular os diversos códigos morais alternativos.
- (2) Determinar o valor das consequências da aceitação colectiva de cada um desses códigos.
- (3) Escolher o código cujas consequências sejam melhores do que as consequências dos outros códigos — é esse o código ideal.

Epistemicamente, este procedimento é demasiado exigente. Jamais conseguiríamos chegar a (3). Visto que os códigos morais alternativos são, em rigor, infinitos, a execução de (1) não é possível. Suponha-se, no entanto, que encontrávamos uma forma de identificar, no conjunto de todos os códigos possíveis, o subconjunto dos códigos que fossem candidatos plausíveis a código ideal — e suponha-se também que esse subconjunto não era intoleravelmente numeroso. Ainda assim, chegados a (2), esbarraríamos em dificuldades formidáveis: determinar o valor das consequências desses códigos no seu todo, bem



como ponderar os seus custos de interiorização, exigiria um conhecimento empírico que pura e simplesmente não está ao nosso alcance. Se não existisse uma alternativa ao procedimento acima delineado, o consequencialista das regras teria de admitir que o código ideal é inescrutável. Isso tornaria a sua teoria muito pouco apelativa. De acordo com um requisito que se impõe se entendermos que cabe às teorias da obrigação oferecer orientação prática, estas têm de ser suficientemente *aplicáveis* — aplicáveis por agentes humanos, com as suas limitações cognitivas. Ora, se o código ideal for inescrutável, e assim ficarmos às escuras quanto ao conteúdo dos princípios optimíficos, o consequencialismo das regras não satisfará este requisito.

Para obter conhecimento dos princípios optimíficos, no entanto, não precisamos de formular o código ideal integralmente e de comparar as suas consequências relevantes com as consequências relevantes de outros códigos morais, também integralmente formulados. Podemos adoptar uma *abordagem parcelar* na tentativa de identificar os verdadeiros princípios optimíficos. Esta abordagem é caracterizável através de dois procedimentos, sendo o primeiro a seguinte regra *prima facie* de especificidade crescente: comecemos por identificar os traços mais gerais do código ideal, avançando depois, gradualmente, para os seus aspectos mais específicos. O segundo é um procedimento de contraste que, restringido a dois códigos, admite esta formulação:

- (1) Consideremos dois códigos,  $C_1$  e  $C_2$ , que (a) têm todas as propriedades já identificadas do código ideal e (b) diferem só a respeito da propriedade  $P$ , de modo que só  $C_1$  tem  $P$ .
- (2) Concentrando-nos nessa diferença, estimemos o valor das consequências da aceitação colectiva de  $C_1$  e de  $C_2$ .

A avaliação das consequências, claro, exigirá reflexão apoiada pela evidência empírica relevante. Se concluirmos que o melhor código é  $C_1$ , isso será uma razão para acreditar que o código ideal tem  $P$ ; se chegarmos à conclusão oposta, teremos uma razão para pensar que o código ideal não tem  $P$ . Nem sempre será fácil — e por vezes poderá revelar-se demasiado difícil — descobrir qual dos códigos contrastados teria as melhores consequências, mas um pessimismo extremo quanto à possibilidade de conseguir conhecimento dos princípios optimíficos seria injustificado.

Na verdade, algum desse conhecimento é fácil de obter. Como vimos na secção anterior, se contrastarmos um código que consiste simplesmente numa exigência ilimitada de maximizar o bem com um código mais leniente, no qual o dever de beneficência esteja limitado por *prerrogativas* para não maximizar o bem, podemos concluir, apelando apenas a um conhecimento elementar da psicologia humana, que o código ideal incluirá essas prerrogativas. Não é difícil ir um pouco mais longe. Sabendo que o código ideal inclui prerrogativas para não maximizar o bem, perguntemos agora se inclui também certas *restrições deontológicas*, como proibições de maltratar e enganar pessoas inocentes. Contrastemos, então, um código com prerrogativas, mas sem restrições, com um código que inclui tanto prerrogativas como restrições. Percebe-se facilmente que, sob a aceitação colectiva de um código como o primeiro, todos seriam demasiado livres para maltratar e enganar inocentes não só para maximizar imparcialmente o bem, mas também por simples interesse pessoal. Isto dá-nos uma razão para crer que o código ideal, além de prerrogativas, inclui restrições. Mas que prerrogativas? E que restrições? Para estas questões mais específicas, e a muitas outras de especificidade ainda maior, poderemos também tentar responder contrastando códigos que difiram apenas no aspecto sob escrutínio.

Dadas as considerações precedentes, não é de estranhar que os consequencialistas das regras defendam que a sua teoria conduz a uma *ética deontológica*. Numa ética deste género, normalmente admite-se — e os consequencialistas das regras admitem — um dever de beneficência geral. Contudo, entende-se que este dever está limitado por prerrogativas (que tornam opcional, e assim permissível, não maximizar o bem em certas circunstâncias) e por restrições (que, correspondendo a certos direitos morais, tornam errado ou impermissível maximizar o bem de algumas formas). À semelhança de muitos outros deontologistas, os consequencialistas das regras afirmam ainda a existência de *obrigações especiais* (como deveres de reparação, deveres parentais ou o dever de cumprir promessas), que são obrigações que alguns indivíduos têm para com outros em virtude de manterem com eles certas relações eticamente significativas.

A preocupação com a acessibilidade epistémica dos princípios optimíficos levou Hooker (2000: 72-75) a propor uma versão expectabilista do consequencialismo das regras. A seu ver, identificar os

princípios que teriam *efectivamente* as melhores consequências é demasiado difícil. Por isso, defende, há que conceber o código ideal não como aquele que tem o maior valor efectivo ou actual, mas como aquele que tem o maior valor esperado. Supostamente, determinar o valor esperado de um código moral será mais fácil do que determinar o seu valor efectivo. Hooker delinea deste modo o procedimento de determinação do valor esperado de um código moral:

A aceitação de um conjunto de regras tem, como é óbvio, vários resultados alternativos possíveis. Suponha-se que podemos identificar o valor ou o desvalor de cada resultado possível. Multiplique-se o valor de cada resultado possível pela probabilidade de esse resultado ocorrer. Adicionem-se todos os produtos dessas multiplicações. O número resultante é o valor esperado desse conjunto de regras. (2016: Secção 6.1)

Este procedimento, na verdade, afigura-se bastante complexo: identificar os vários resultados possíveis da aceitação colectiva de um código, bem como estimar o valor e a probabilidade de cada um desses resultados, seria uma tarefa intimidante. Repeti-la para todos os candidatos a código ideal, até encontrarmos o vencedor, seria tudo menos simples.<sup>2</sup> Mas, como vimos, em vez de avaliar códigos alternativos integralmente formulados, o melhor será optar por uma abordagem parcelar na tentativa de identificar os princípios optimíficos. Adoptada esta abordagem, teremos razões para preferir um consequencialismo das regras expectabilista a uma versão actualista da teoria? Dificilmente. Não é claro que estimar o valor efectivo ou actual de princípios morais, um por um, contrastando-os com princípios alternativos, seja mais difícil do que estimar o seu valor esperado.

#### 4 Níveis de aceitação

De uma forma propositadamente vaga, o consequencialismo das regras foi apresentado como a teoria de que devemos agir segundo os princípios cuja aceitação *colectiva* resultaria em melhores consequências. Sob a interpretação mais natural, a aceitação colectiva cor-

<sup>2</sup> Veja-se Galvão (2008: 202-208).

responde a aceitação universal: devemos agir segundo os princípios que maximizariam o bem se todos os agentes racionais os aceitassem. No entanto, Brad Hooker (2000: 80-85) rejeita esta versão de consequencialismo das regras. Ao avaliar princípios morais pensando numa situação de aceitação universal, diz-nos, estaremos a avaliá-los tendo em mente um cenário demasiado idealizado. Esse é um cenário que exclui tanto agentes com ideias morais profundamente erradas (chamemos-lhes *fanáticos*) como agentes empedernidamente amoralistas. Por isso, se escolhermos o código que teria as melhores consequências se fosse aceite universalmente, estaremos a escolher um código desfasado da realidade — um código que não nos dará os princípios adequados para lidar com fanáticos e amoralistas.

Qual é a alternativa, então? Para Hooker, há que conceber o código ideal como aquele que maximizaria o bem se fosse aceite pela grande maioria dos agentes, mas não por todos eles. Hooker chega mesmo a precisar a dimensão dessa maioria, sugerindo que pensemos num cenário de aceitação por 90% dos agentes. A referência a uma ‘grande maioria’ é bastante vaga, mas a percentagem indicada é, sem dúvida, bastante arbitrária. Poderá parecer-nos, razoavelmente, que tanto a vagueza como a arbitrariedade em causa são indesejáveis, ou mesmo intoleráveis, na formulação de um princípio ético que se apresenta como fundamental.

Michael Ridge (Ridge 2006) apresenta uma alternativa tanto à perspectiva da aceitação universal como à perspectiva da aceitação geral de Hooker. Estas perspectivas, diz-nos, assemelham-se no seguinte aspecto: pressupõem que o código ideal é aquele que teria melhores consequências sob um ‘índice fixo’ de aceitação. Ridge rejeita este pressuposto, optando por um consequencialismo das regras de ‘índice variável’, cujo conteúdo é elucidado na seguinte passagem:

Em vez de privilegiarmos um nível específico de aceitação social, poderemos, na verdade, incluir todos os níveis de aceitação social possíveis na explicação do que é um acto correcto. Mais precisamente, poderemos sustentar que um acto é correcto se, e somente se, é exigido pelas regras que têm a seguinte propriedade: quando se determina a utilidade esperada de cada nível de aceitação entre (e incluindo) 0% e 100% para essas regras, e se calcula a utilidade esperada média para todos esses níveis de aceitação

diferentes, a média para essas regras é pelo menos tão elevada como a média correspondente a cada conjunto de regras alternativo. (2006: 248)

Para ilustrar a perspectiva de Ridge, consideremos dois códigos morais,  $C_1$  e  $C_2$ , cujo valor em certos níveis de aceitação possíveis é o indicado na tabela seguinte:

	...32%	33%	34%	35%...
$C_1$	$n$	$n + 1$	$n + 1$	$n + 2$
$C_2$	$n - 6$	$n - 4$	$n - 2$	$n$

Imagine-se que conhecíamos também o valor das consequências de  $C_1$  e de  $C_2$  para cada um dos restantes noventa e sete níveis de aceitação possíveis, e que depois, para cada código, calculávamos o seu valor médio.  $C_1$  seria melhor do que  $C_2$  se, e somente se, tivesse um valor médio superior a  $C_2$ . E o código ideal seria aquele que, comparado com os códigos alternativos, teria o valor médio mais elevado.

A proposta de Ridge não prima pela elegância. Tem um defeito mais significativo, que é o de tornar mais difícil — porventura demasiado difícil — a identificação dos princípios optimíficos, na medida em que implica avaliar princípios à luz de uma grande quantidade de níveis de aceitação possíveis.<sup>3</sup>

Não é claro, pois, que as perspectivas de Hooker ou de Ridge sejam preferíveis à versão de consequencialismo das regras mais elegante, na qual aquilo que conta são as consequências numa situação de aceitação universal. Importa recordar que a aceitação de princípios não implica o seu estrito cumprimento. Um agente pode aceitar um princípio moral mas não o cumprir porque, por falta de conhecimento, o aplicou mal, ou então porque o interesse pessoal se sobrepôs à sua motivação para agir moralmente. Assim, quando consideramos uma situação de aceitação universal, não estamos a imaginar um cenário no qual engano moral e o amoralismo estão ausentes. Talvez esse cenário, então, não seja excessivamente idealizado para escolher princípios adequados ao mundo real.

<sup>3</sup> Para uma crítica a Ridge, veja-se Galvão (2008: 198-201).

## 5 Relativização do código ideal?

Será o código ideal o mesmo para todos os agentes? Ou, pelo contrário, o código ideal para alguns agentes será diferente do código ideal para outros agentes? Restringindo a sua atenção a agentes humanos, Brad Hooker (2000: 85-88) recusa-se a relativizar o código ideal, defendendo assim uma versão de consequencialismo das regras que — com uma qualificação, como veremos — afirma a existência de um único código ideal para todos os agentes de todas as gerações e sociedades.

Uma hipótese seria relativizar o código ideal a grupos sociais. Poderia haver, por exemplo, um código para os mais inteligentes e outro para os menos inteligentes, um código para as mulheres e outro para os homens. Opondo-se a esta hipótese, Hooker afirma o seguinte:

[A] ideia de relativizar códigos a grupos é uma via para relativizá-los a subgrupos, e no fim dessa via encontra-se a sua relativização a indivíduos. Seguir esta via é virar as costas a um dos aspectos tradicionalmente atraentes do consequencialismo das regras — nomeadamente, o facto de se basear na ideia de que a moralidade deve ser concebida como um código *colectivo e partilhado*. (2000: 87)

Note-se que Hooker não defende que o código ideal terá de ser o mesmo para todos os agentes racionais concebíveis. Se existirem extraterrestres racionais com traços psicológicos muito diferentes dos humanos, os princípios optimíficos para esses agentes poderão divergir profundamente dos princípios optimíficos para os agentes humanos. Mas suponha-se agora que, com o uso de tecnologias de aperfeiçoamento genético, as capacidades cognitivas e motivacionais dos seres humanos se alteram muito significativamente. Considerando uma suposição como esta, Hooker (2005: 268-269) acaba por sustentar que o código ideal para nós é aquele maximizaria o bem se fosse interiorizado pela grande maioria das pessoas — em todos os lugares e em cada nova geração, *excluindo as gerações posteriores ao desenvolvimento de tecnologias que reduzam significativamente o custo de*

*interiorização das regras morais.* O código ideal para os agentes destas gerações, então, poderá ser substancialmente diferente daquele que se nos adequa.

Uma alternativa melhor será ir mais longe e relativizar o código ideal a sociedades, como Brandt (1979: 300) sugere. Este tipo de relativização, ao invés da relativização a grupos coexistentes numa sociedade, não põe em causa a concepção da moralidade em termos de um código ‘colectivo e partilhado’. Além do mais, afigura-se implausível que os princípios optimíficos para nós, agentes humanos do século XXI, sejam iguais aos princípios optimíficos para os habitantes da Suméria do século XXV a.C., digamos. Dado que partilhamos com os sumérios os traços da natureza humana, podemos admitir que, nos seus aspectos mais gerais, o código ideal para nós coincidirá com o código ideal para eles e para qualquer outra sociedade humana que já existiu. No entanto, as sociedades humanas enfrentam problemas muito diferentes e dispõem de tecnologias diferentes, pelo que, se advogarmos um único código ideal para toda a humanidade, recusando-nos a relativizá-lo a sociedades, teremos de concluir que esse código é insensível a todas essas diferenças. Esta não é uma via promissora para o consequencialismo das regras.

## 6 Harsanyi: uma justificação contratualista

Por que razão haveremos de aceitar o consequencialismo das regras? Para justificar esta teoria, podemos tentar mostrar que ela decorre de uma perspectiva ética com maior plausibilidade inicial. Uma hipótese será partirmos do contratualismo ético, que afirma que os princípios morais correctos são aqueles que seriam escolhidos consensualmente, sendo assim acordados, por agentes racionais. Na versão rawlsiana de contratualismo, os agentes escolhem princípios para a sociedade na qual hão-de viver, mas, sujeitos a um véu de ignorância, ignoram o lugar que irão ter nessa sociedade. Desconhecem, por exemplo, o seu sexo, a sua raça ou a sua religião. Supõe-se que esta ignorância tornará a escolha apropriadamente imparcial.

John Rawls (Rawls 1971) desenvolveu uma teoria da justiça na qual este género de contratualismo resulta em princípios adversos ao utilitarismo. Antes de Rawls, Harsanyi (Harsanyi 1953, 1955, 1977) apresentou um contratualismo rawlsiano — só que em defesa do

utilitarismo. Mais precisamente, Harsanyi defendeu um utilitarismo no qual se concebe o bem-estar como satisfação de preferências. Examinemos o seu argumento.

### 6.1 O argumento da equiprobabilidade

Os agentes de Harsanyi, além de estarem sob um véu de ignorância, são racionais no seguinte sentido: escolhem aquilo que maximiza a utilidade esperada, isto é, aquilo que expectavelmente resultará na maior satisfação das suas preferências. Imaginemos, como sugere Harsanyi, que lhes cabe escolher entre diversos ‘sistemas sociais’. Dado que estes agentes não sabem que lugar terão na sociedade, atribuem a mesma probabilidade a cada um dos lugares possíveis nessa sociedade. Harsanyi designa este pressuposto por *postulado da equiprobabilidade*.

Através de um exemplo muito similar ao de Harsanyi (1977: 44-46), vejamos como a sua perspectiva conduz ao utilitarismo. Imaginando que um destes agentes está a escolher entre dois sistemas sociais, uma democracia e uma ditadura, suponha-se que a sociedade a que essa escolha diz respeito consiste em  $n$  indivíduos. Denotemos os níveis de bem-estar, ou de utilidade, de que os indivíduos 1, 2 ...  $n$  fruiriam no sistema social em causa da seguinte maneira:  $U_1, U_2 \dots U_n$ . Sob o véu de ignorância de Harsanyi, o agente atribuirá a mesma probabilidade,  $1/n$ , ao resultado de ocupar cada posição social e, consequentemente, à fruição de cada um dos níveis de bem-estar  $U_1, U_2 \dots U_n$ . Sendo racional, o agente escolherá o sistema social que lhe ofereça as melhores perspectivas de bem-estar. Como poderá fazer isso? Para simplificar, suponha-se que temos apenas quatro cidadãos e que os níveis de bem-estar em cada sistema são os seguintes:

Democracia	$U_1 = 4$	$U_2 = 3$	$U_3 = 3$	$U_4 = 2$
Ditadura	$U_1 = 6$	$U_2 = 2$	$U_3 = 1$	$U_4 = 1$

Como não sabe que posição ocupará nestes sistemas sociais, o agente atribuirá uma probabilidade de  $1/4$  à hipótese de, em cada sistema, ocupar cada uma das posições possíveis. Poderá então calcular a utilidade esperada de ambas as opções. Os cálculos para a opção da



democracia e para a opção da ditadura são, respectivamente, os seguintes:

$$\begin{aligned}(4 \frac{1}{4}) + (3 \frac{1}{4}) + (3 \frac{1}{4}) + (2 \frac{1}{4}) &= 3 \\ (6 \frac{1}{4}) + (2 \frac{1}{4}) + (1 \frac{1}{4}) + (1 \frac{1}{4}) &= 2.5\end{aligned}$$

Interessado em maximizar a utilidade esperada, o Decisor escolherá, nestas circunstâncias, o sistema democrático. Ao fazê-lo, terá escolhido o sistema que, considerados todos os indivíduos, exhibe o *maior bem-estar médio*. Como podemos constatar, o bem-estar médio na democracia e na ditadura correspondem, respectivamente, a 3 e a 2.5.

Podemos ver agora que resultado Harsanyi pretende estabelecer. Os agentes racionais, em seu entender, maximizam a utilidade esperada. Sob um véu de ignorância que os leva a assumir um ponto de vista moral, escolhem as opções que maximizam o bem-estar médio de toda a população considerada, isto é, adoptam um utilitarismo de média. Admitindo que cada sistema social se caracteriza pela aceitação de um determinado código moral, podemos concluir que os agentes racionais escolherão unanimemente o código cuja aceitação resulte no maior bem-estar médio.

## 6.2 Críticas ao argumento da equiprobabilidade

Rawls conta-se entre os críticos de Harsanyi.<sup>4</sup> A sua crítica incide sobretudo no postulado da equiprobabilidade, que, a seu ver, resulta numa versão implausível de contratualismo. Segundo Rawls, há que conceber as partes contratantes de outra forma, colocando-as sob um véu de ignorância mais espesso, que as faz ignorar a probabilidade de ficarem nas diversas posições sociais possíveis. Poderá ocorrer-nos que, dada essa ignorância, o mais sensato seria aplicar a regra da razão insuficiente e assim, como Harsanyi recomenda, atribuir probabilidades iguais aos vários resultados possíveis. Rawls discorda. Para si, o princípio de escolha apropriado para as partes contratantes é a regra *maximin*, que nos diz para escolher a opção cujo pior resultado possível seja o melhor. Adoptando esta regra, os agentes irão supor

<sup>4</sup> Para o debate entre Rawls e Harsanyi, veja-se Rawls (1974) e Harsanyi (1975). Veja-se também Parfit (2011: 346-354).

que ficarão na pior posição social, pelo que escolherão o sistema social no qual aqueles que ocupam essa posição estão, ainda assim, menos mal.

Contra Rawls, Harsanyi observa que a regra maximin só ilusoriamente dispensa o recurso a probabilidades. Quem a emprega, na verdade, está a atribuir tacitamente uma probabilidade de 100% à ocorrência do pior resultado possível — uma probabilidade insensata, que corresponde a uma atitude extremamente pessimista.

Há objecções melhores ao argumento de Harsanyi. Podemos alegar que o seu resultado, o utilitarismo de média, é uma versão particularmente implausível de consequencialismo. Quando comparamos situações com populações que têm a mesma dimensão, não fará diferença avaliá-las à luz do bem-estar médio ou simplesmente do bem-estar total. No entanto, se essas comparações envolverem populações com dimensões distintas, a situação preferível em termos de bem-estar médio poderá não ser aquela em que há um maior bem-estar total. Algumas destas avaliações divergentes revelam contundentemente a implausibilidade do utilitarismo de média. Imagine-se, por exemplo, uma população constituída por dez mil milhões de pessoas muito felizes e outra população, com apenas dez pessoas, constituída por pessoas também muito felizes — na verdade, ligeiramente mais felizes do que as outras. Visto que nesta segunda população o bem-estar médio é mais elevado, o utilitarista de média terá de considerá-la preferível à primeira, o que é bastante contrainutivo. Suponha-se agora que a população mais pequena passa a ter onze pessoas. Esta décima primeira pessoa é também muito feliz, embora não tão feliz como as outras dez. O utilitarista de média terá de dizer, de forma extremamente contra-intuitiva, que a situação *piorou* — que teria sido melhor que essa pessoa nunca tivesse nascido. Em suma, se o argumento da equiprobabilidade resulta num consequencialismo das regras em que o código ideal visa o maior bem-estar médio, o reconhecimento de que o bem-estar médio não é algo a visar fundamentalmente torna muito suspeitos os pressupostos do argumento.

Na verdade, como Parfit deixou claro, o contratualismo rawlsiano sofre de falhas irremediáveis em questões de ética populacional. Suponha-se que estamos a escolher entre princípios cuja aceitação resulta em populações com dimensões diferentes. Será apropriado

fazer essa escolha em função dos nossos interesses, sob um véu de ignorância? Ao fazê-la, sabendo que nós próprios existimos, estaríamos a presumir que existiremos, seja qual for o princípio escolhido. Parfit considera que esta suposição é ilícita:

Este pressuposto viola o requisito de imparcialidade. O princípio que escolhermos afectará a quantidade de pessoas existentes. Se presumimos que existiremos de certeza, seja qual for o princípio que escolhamos, isso é como presumir, quando escolhemos um princípio que prejudicará as mulheres, que seremos homens de certeza. (1984: 392)

Mas conceda-se a licitude de supor que, seja qual for o princípio escolhido, existiremos de certeza. A partir dos casos seguintes, Parfit argumenta que, ainda assim, o contratualismo rawlsiano não é defensável:

No Inferno Um, a última geração consiste em dez pessoas inocentes. Cada uma delas fica em grande agonia durante cinquenta anos. As vidas destas pessoas são muito piores do que nada. Se pudessem, elas cometeriam suicídio.

No Inferno Dois, a última geração consiste não em dez, mas em dez milhões de pessoas inocentes. Cada uma delas fica numa agonia igual durante cinquenta anos menos um dia. (1984: 393)

De acordo com o contratualismo rawlsiano, a escolha correcta seria o princípio conducente ao Inferno Dois — o que é manifestamente contra-intuitivo, para não dizer absurdo.

Pelo menos enquanto teoria geral da obrigação, então, o contratualismo no qual Harsanyi baseia a sua defesa do consequencialismo das regras não é uma perspectiva promissora. Como vimos, pode dizer-se o mesmo da versão de consequencialismo que resulta do argumento da equiprobabilidade.

## 7 Parfit: uma justificação kantiana

Deve-se a Parfit outra tentativa de fundamentar o consequencialismo das regras numa perspectiva ética distinta. Mais precisamente, Parfit procura mostrar que a Fórmula da Lei Universal do imperativo categórico kantiano, devidamente corrigida, conduz ao consequencialismo

das regras. O seu *Argumento Kantiano a favor do Consequencialismo das Regras* é o seguinte:

- (A) Todos devem seguir os princípios cuja aceitação universal todos possam querer ou escolher racionalmente.
  - (B) Todos podem escolher racionalmente algo se têm razões suficientes para o escolher.
  - (C) Há princípios AU-optimíficos [i.e. princípios cuja aceitação universal resultaria nas melhores consequências].
  - (D) Estes são os princípios que todos teriam as razões imparciais mais fortes para escolher.
  - (E) As razões imparciais de alguém para escolher estes princípios nunca são decisivamente suplantadas por quaisquer razões relevantes que estejam em conflito com estas.
- Logo
- (F) Todos têm razões suficientes para escolher os princípios optimíficos.
  - (G) Não há princípios significativamente não-optimíficos que todos tenham razões suficientes para escolher.
- Logo
- (H) Os únicos princípios que todos têm razões suficientes para escolher, e que assim poderão escolher racionalmente, são os princípios optimíficos.
- Logo, todos devem seguir estes princípios. (2011a: 378)

Em suma, a injunção kantiana de agir segundo princípios que, racionalmente, possamos querer como leis universais leva, segundo Parfit, à injunção de seguir os princípios optimíficos.

Começemos por examinar os elementos dificilmente controversos do Argumento Kantiano, que são as premissas (B) e (C). A primeira destas premissas diz-nos apenas que, para podermos escolher algo racionalmente, basta termos razões *suficientes* para o fazermos. Essas razões não têm de ser *decisivas*. Suponha-se que temos razões suficientes para escolher duas de cinco alternativas disponí-

veis. As razões para escolher uma delas não suplantam decisivamente as razões para escolher a outra. Nesse caso, poderemos escolher racionalmente qualquer uma dessas duas alternativas, mas não uma das outras três.

A premissa (C) diz-nos apenas que há princípios optimíficos. Não afirma, claro, que esses são os princípios correctos, visto que essa é a conclusão do argumento. Que princípios contarão como optimíficos, isso dependerá em parte da concepção do bem que o consequencialista das regras subscreva.

### 7.1 Razões práticas

O Argumento Kantiano não está comprometido com uma concepção definida do bem. No entanto, assenta numa concepção bastante definida da natureza das razões práticas, ou seja, das razões (normativas) para agir, querer ou escolher. Parfit rejeita todas as versões de subjectivismo das razões, isto é, todas as perspectivas segundo as quais as razões de um agente resultam de algum modo dos seus próprios desejos, fins ou escolhas. Em seu entender, todas estas perspectivas implicam, absurdamente, que um agente *pode* não ter a menor razão para evitar a sua agonia futura, na medida em que a existência dessa razão dependerá dos desejos que ele tem ou teria, ou das escolhas que ele faria.<sup>5</sup>

Parfit advoga antes um objectivismo das razões: as razões práticas decorrem não do sujeito, mas do valor (ou desvalor) que certas coisas têm, independentemente das atitudes de qualquer sujeito. A agonia, por exemplo, é má pela sua própria natureza, pelo que qualquer agente tem sempre uma razão para evitar a sua agonia futura — ainda que, em alguns casos, possa ter razões mais fortes para suportar a agonia.

Pode-se ser objectivista das razões de maneiras muito diferentes. Parfit distingue — e repudia — duas versões extremas de objectivismo:

*Egoísmo Racional*: Temos sempre razões mais fortes para fazer aquilo que seria melhor para nós próprios.

(...)

<sup>5</sup> Para uma crítica a Parfit, veja-se Galvão (2018).

*Imparcialismo Racional:* Temos sempre razões mais fortes para fazer aquilo que seria melhor imparcialmente. (2011a:130)

Suponha-se que salvar várias pessoas resultaria num ligeiríssimo incómodo para um determinado agente, sem que ele colhesse qualquer benefício do salvamento. A perspectiva egoísta implica, implausivelmente, que seria irracional esse agente suportar o incómodo para salvar essas pessoas. Suponha-se agora que, fazendo um sacrifício pessoal tremendo, um certo agente tornaria as coisas ligeiramente melhores de uma perspectiva imparcial. Também implausivelmente, a perspectiva imparcialista implica que seria irracional esse agente não fazer o sacrifício.

Até certo ponto inspirado por Henry Sidgwick (1907), Parfit opta por uma posição intermédia, que formula deste modo:

*Perspectivas objectivas amplas baseadas no valor:* Quando um de dois actos possíveis faria as coisas correrem de uma forma imparcialmente melhor, mas o outro acto faria as coisas correrem melhor para nós mesmos ou para aqueles com os quais mantemos laços estreitos, frequentemente temos razões suficientes para agir de qualquer uma dessas formas. (2011a: 137)

As perspectivas deste género são amplas porque, ao invés do egoísmo e do imparcialismo, reconhecem tanto razões parciais (ou pessoais) como imparciais. As primeiras, note-se, incluem não só razões ditadas pelo interesse pessoal, mas também razões para fazer aquilo que será melhor para aqueles que nos são mais próximos.

Parfit acrescenta que as razões parciais e imparciais, embora não sejam incomensuráveis, podem ser comparadas só de uma forma muito imprecisa. Ainda assim, em alguns casos será claro que umas suplantam as outras. Daí que, quando estas apontam para acções incompatíveis, só frequentemente tenhamos razões suficientes para fazer tanto aquilo que será melhor imparcialmente como aquilo que será melhor para nós próprios ou para aqueles que nos são mais próximos. Por vezes, como nos casos acima delineados, isso não acontece. Presumivelmente, Parfit diria que teríamos uma razão decisiva para salvar várias pessoas à custa de um ligeiríssimo incómodo, bem como uma razão decisiva para não fazer um sacrifício tremendo de modo a tornar as coisas apenas ligeiramente melhores de uma

perspectiva imparcial.

### 7.2 *Premissa (A)*

Retomando a análise do Argumento Kantiano, recordemos a premissa (A):

- (A) Todos devem seguir os princípios cuja aceitação universal todos possam querer ou escolher racionalmente.

Esta premissa, que Parfit designa por *contratualismo kantiano*, é o resultado de uma longa discussão da Fórmula da Lei Universal do imperativo categórico de Kant (1785), que nos diz o seguinte numa das suas versões mais conhecidas:

- (K) Age apenas segundo uma máxima tal que possas ao mesmo tempo querer que ela se torne lei universal.

Além de dispensar o conceito kantiano de máxima, Parfit, em (A), faz outra revisão significativa de (K), que podemos captar contrastando os imperativos seguintes:

- (A\*) Segue os princípios cuja aceitação universal *todos* possam querer ou escolher racionalmente.  
(K\*) Segue os princípios cuja aceitação universal *tu* possas querer ou escolher racionalmente.

(K\*) está mais próximo da Fórmula da Lei Universal. Segundo Parfit (2011a: 334-338), este princípio tem implicações inaceitáveis. Por exemplo, um racista pode querer racionalmente princípios de segregação racial, visto que poderá ter razões suficientes, parciais, para querer que esses princípios sejam aceites universalmente. Do mesmo modo, um homem pode ter razões suficientes, parciais, para querer que todos aceitem princípios de discriminação das mulheres, e um rico pode ter razões suficientes, uma vez mais parciais, para querer que todos aceitem princípios muito adversos aos pobres.

(A\*) — a formulação imperativa de (A) — não tem implicações como estas. Ainda que o racista tenha razões suficientes para querer que todos aceitem princípios de segregação racial, aqueles que sofrem com essa segregação não têm razões suficientes, nem parciais nem imparciais, para querer a aceitação universal desses princípios. Portanto,

(A\*) condena princípios de segregação racial. Podemos dizer algo semelhante a respeito dos outros dois casos.

É por (A) se referir ao que *todos* podem querer racionalmente como lei universal que Parfit designa esta premissa por contratualismo kantiano. Mas existirão princípios que todos os agentes possam querer deste modo? Quem aceite um subjectivismo das razões ou o egoísmo racional dificilmente poderá dar uma resposta afirmativa a esta questão. Como vimos, no entanto, Parfit rejeita estas perspectivas sobre as razões normativas.

### 7.3 *Premissa (D)*

De acordo com (C), há princípios optimíficos. A premissa (D) limita-se a identificá-los com ‘os princípios que todos teriam as razões imparciais mais fortes para escolher’. Esta identificação poderá parecer trivial, mas, como Michael Otsuka (2009: 63-69) argumenta, admite contestação.

A identificação será correcta só se as razões imparciais forem, todas elas, razões para *promover* certos valores (e.g. a felicidade, a igualdade de felicidade) tanto quanto possível, gerando assim as melhores consequências, vistas as coisas de uma perspectiva impessoal. No entanto, talvez existam certos valores que exigem uma atitude de *respeito*, e não de promoção. Por exemplo, um kantiano poderá alegar que a autonomia, ou a natureza racional dos agentes, é um valor a respeitar, não a promover, de tal modo que seria inaceitável sacrificar a autonomia de alguns para benefício da autonomia de outros. E o kantiano poderá acrescentar, plausivelmente, que as razões para respeitar a autonomia não são razões parciais, mas imparciais, recusando-se assim a identificar os princípios que ‘todos teriam as razões imparciais mais fortes para escolher’ com os princípios cuja aceitação universal mais promoveria o bem.

### 7.4 *Premissa (E)*

Recordemos agora a premissa (E), que afirma o seguinte a respeito dos princípios optimíficos:

(E) As razões imparciais de alguém para escolher estes princípios



nunca são decisivamente suplantadas por quaisquer razões relevantes que estejam em conflito com estas.

Parfit (2011a: 380-398) empreende uma defesa bastante minuciosa desta premissa do Argumento Kantiano. Detenhamo-nos apenas em alguns dos seus aspectos, que envolvem casos hipotéticos de salvamentos com baleeiras. Parfit descreve assim o primeiro desses casos:

[E]stou preso num rochedo e cinco pessoas estão presas noutra rochedo. Antes que a maré suba ao ponto de cobrir ambos os rochedos, um agente poderá usar uma baleeira para me salvar ou para salvar as outras cinco pessoas. Eu e essas pessoas somos estranhos tanto para o agente como entre nós. Somos relevantemente semelhantes de outras formas. Todos somos jovens e, se morrermos, todos perderemos muitos anos de vida feliz. (2011a: 380)

Um princípio optimífico, afirma Parfit, dirá seguramente que o agente deve salvar as cinco pessoas. A título de exemplo, Parfit propõe o princípio seguinte:

*Princípio da Quantidade:* Quando podemos salvar um de dois grupos de pessoas, sendo todas elas estranhas para nós e relevantemente semelhantes de outras formas, devemos salvar o grupo que inclui mais pessoas. (2011a: 380)

Supondo agora que Parfit está no rochedo mais próximo do agente, consideremos um princípio alternativo, que não é optimífico:

*Princípio da Proximidade:* Nesses casos, devemos salvar o grupo que esteja mais próximo de nós. (2011a: 381)

Parfit, admitamos, tem razões imparciais para escolher o Princípio da Quantidade em vez do Princípio da Proximidade. Mas, dadas as circunstâncias, não serão essas razões decisivamente suplantadas por razões parciais ditadas pelo interesse pessoal, revelando-se assim a falsidade de (E)? Afinal, se o agente da baleeira seguir o Princípio da Quantidade, cinco pessoas serão salvas, mas Parfit morrerá.

Baseando-se na sua teoria das razões, Parfit responde negativamente. No caso em apreço, diz-nos, teria razões suficientes para escolher que o agente o salvasse, mas também teria razões suficientes

para escolher antes que ele salvasse as cinco pessoas, pelo que poderia escolher racionalmente que todos aceitassem o Princípio da Quantidade em vez do Princípio da Proximidade.

Consideremos agora um caso semelhante, que poderá afigurar-se mais ameaçador para o Argumento Kantiano. Agora somos nós quem está na baleeira – e podemos salvar o nosso próprio filho ou cinco estranhos. A respeito deste caso, Parfit declara o seguinte:

Podemos acreditar que, mesmo que possamos abdicar racionalmente da *nossa* vida para salvar cinco estranhos, não poderemos abdicar racionalmente da vida do nosso *filho* para salvar esses estranhos, nem poderemos escolher racionalmente que todos aceitem um princípio optimífico que exija esse acto. Parece, então, que este caso oferece uma objecção à premissa (E). (2011a: 385)

No entanto, alega Parfit, não é verdade que os princípios optimíficos exigissem salvar os estranhos. Muito plausivelmente, as coisas correrão melhor, de uma perspectiva imparcial, sob a aceitação universal de princípios que autorizem os pais a dar uma forte prioridade aos interesses dos *seus* filhos.

#### 7.5 *Premissa (G)*

Para benefício da discussão, admitamos que Parfit defende convincentemente a premissa (E) do Argumento Kantiano. Do seu ponto de vista, a premissa (G) é muito menos problemática. Recordemo-la:

(G) Não há princípios significativamente não-optimíficos que todos tenham razões suficientes para escolher.

Eis a justificação que Parfit oferece para (G):

Se todos aceitassem [princípios significativamente não-optimíficos], as coisas correriam de formas que seriam imparcialmente muito piores. Dizer que esses princípios são significativamente não-optimíficos significa isto mesmo. Esses factos [sobre como as coisas correriam] dariam a todos fortes razões imparciais para não escolherem que todos aceitassem tais princípios. Visto que a maioria das pessoas não teria razões pessoais em conflito com estas, a maioria das pessoas não poderia fazer racionalmente essa

escolha. E, em praticamente todos esses casos, *se todos aceitassem um desses princípios não-optimíficos, as coisas correriam muito pior também para algumas pessoas infortunadas*. É ainda mais claro que essas pessoas não poderiam escolher racionalmente que todos aceitassem esse princípio, dado que essas pessoas teriam tanto fortes razões imparciais como fortes razões pessoais para não fazer essa escolha. (2011a: 399-400, ênfase aduzido)

Será esta justificação de (G) satisfatória? Segundo Jacob Ross (2009), não. Ross delinea alguns contra-exemplos a (G), isto é, algumas situações possíveis nas quais **há** princípios significativamente não-optimíficos que, à luz da teoria das razões que Parfit subscreve, todos têm razões suficientes para escolher.

O melhor dos contra-exemplos de Ross consiste, um tanto ironicamente, numa ligeiríssima variante de um caso hipotético que Parfit (1984: 352-379) utilizou para colocar o problema da não-identidade. Imagine-se que estamos a definir políticas energéticas e que os princípios disponíveis são apenas os seguintes:

(FR) Devemos extrair energia de fontes renováveis.

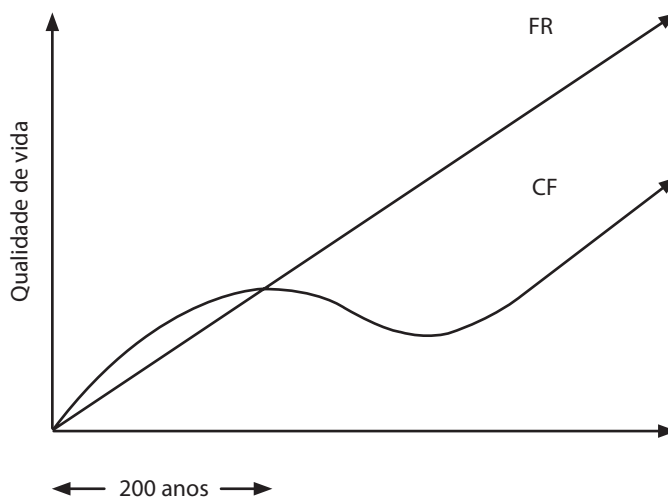
(CF) Devemos extrair energia de combustíveis fósseis até estes se esgotarem.

Considerando este cenário, introduzamos agora várias suposições:

- (1) Em virtude dos seus efeitos ambientais, a aceitação universal de (FR) resultaria, globalmente, em melhores consequências do que a aceitação universal de (CF).
- (2) Durante os próximos duzentos anos, no entanto, todos viverão melhor caso (CF) seja aceite universalmente.
- (3) A escolha entre (FR) e (CF) terá consequências muito vastas, que abrangem a identidade das pessoas futuras. Caso (FR) seja aceite universalmente, ao longo de todo o período da história humana *posterior* aos próximos duzentos anos existirão certas pessoas. Caso (CF) seja aceite, ao longo de todo esse período existirão *outras* pessoas. Ou seja, decorridos duzentos anos, nenhuma pessoa existirá em ambos os cenários.
- (4) Caso todos aceitem (CF), as pessoas que existirem daqui

a mais de duzentos anos terão uma qualidade de vida positiva. Todavia, se todos aceitarem antes (FR), as pessoas que existirem ao longo desse período terão uma qualidade de vida ainda melhor.

O gráfico seguinte ilustra estas suposições:



Dada a suposição (1), (FR) é o princípio optimífico. No entanto, todos têm razões suficientes para preferir a aceitação universal de (CF). Manifestamente, aqueles que existem agora ou que existirão nos próximos duzentos anos têm razões parciais ou pessoais para fazerem essa escolha. E, o que é mais surpreendente, podemos dizer o mesmo daqueles que existirão daqui a mais de duzentos anos: essas pessoas futuras têm uma qualidade de vida positiva e nem sequer existiriam se os seus antepassados tivessem adoptado (FR). Temos aqui, portanto, um princípio significativamente não-optimífico que todos podem escolher racionalmente.

### 7.6 A Teoria Tripla

Além de pressupor uma teoria das razões controversa, o Argumento Kantiano, como acabámos de ver, não é inteiramente convincente. Vale a pena acrescentar que o projecto de Parfit não é apenas o de conciliar as melhores versões da ética kantiana e da ética consequencialista. Este projecto conciliador estende-se à ética contratualista, do que resulta a *Teoria Tripla*. De acordo com esta teoria, os princípios cuja aceitação universal todos podem querer racionalmente coincidem não só com os princípios optimíficos, mas também com os princípios que *ninguém* pode rejeitar razoavelmente num acordo ético. O *contratualismo scanloniano* (Scanlon 1998) diz-nos que estes últimos são os princípios éticos correctos. Embora Parfit inclua esta perspectiva na sua Teoria Tripla, não a defende na mesma versão que o seu autor, T. M. Scanlon.<sup>6</sup>

## 8 Equilíbrio reflexivo

Mesmo que o consequencialismo das regras não seja justificável a partir de outra perspectiva ética, à qual se atribua uma maior plausibilidade inicial, poderão existir razões suficientemente fortes para o aceitarmos. Uma razão que Hooker (2000: 9-30) enfatiza é esta: o consequencialismo das regras oferece uma boa sistematização das nossas convicções ou intuições morais ponderadas. Aceitamos intuitivamente uma série de princípios morais, correspondentes, por exemplo, a deveres de beneficência e não-maleficência, lealdade e veracidade, gratidão e reparação. Uma hipótese — associada a David Ross (1930), e por isso conhecida por *pluralismo rossiano* — será todos esses princípios serem básicos, o que nos deixará com uma pilha desconexa de deveres. Com o consequencialismo das regras, alega Hooker, conseguimos melhor do que isso, pois torna-se claro que esses princípios intuitivamente apelativos decorrem de um só princípio básico — o de que devemos seguir os princípios optimíficos. Em defesa desta posição, Hooker declara o seguinte:

Globalmente, as consequências serão muito melhores se existirem regras aceites pela generalidade das pessoas que proibam o ataque

<sup>6</sup> Parfit discute com Scanlon o contratualismo scanloniano em Parfit (2011b).

físico, a tortura, o roubo, o incumprimento de promessas, a mentira e outras coisas similares. Na verdade, a existência de pelo menos formas mínimas dessas regras é indispensável à sociedade. Veja-se como as coisas correm onde essas regras geralmente não são aceites e observadas. Nesses lugares, a vida desce ao caos e torna-se sórdida, rude e curta, como Hobbes previu. (...) E, como muitos outros observaram, dada a competição pelos recursos e ‘a inescapável divisão do trabalho e a perene necessidade de cooperação’, as regras que protegem a propriedade e que permitem que as pessoas se comprometam com acções futuras, fazendo promessas, são essenciais. O consequencialismo das regras prescreverá assim regras que ‘protegem as pessoas, a propriedade e as promessas’, para usar a expressão de Hart. (2000: 126-127)

Como muitos outros, Hooker aceita a metodologia do equilíbrio reflexivo. Quem adopta esta metodologia procura o melhor equilíbrio, ou a máxima coerência, entre intuições morais e princípios éticos. Por vezes, revelar-se-á desejável reconsiderar algumas das intuições iniciais; noutros casos, será melhor rejeitar um determinado princípio, substituindo-o por um princípio que corresponda melhor às nossas intuições. Fazendo estes ajustes, chegaremos, idealmente, a um estado de equilíbrio reflexivo: teremos descoberto os princípios éticos correctos, que justificarão de um modo elegante, sistemático, aquelas intuições que merecem ser preservadas.

Para Hooker, então, a busca do equilíbrio reflexivo leva-nos a preferir o consequencialismo das regras a teorias alternativas, como o pluralismo rossiano e o consequencialismo dos actos. Importa observar que muitos têm rejeitado esta última teoria precisamente por entenderem que ela tem implicações demasiado contra-intuitivas. Um consequencialista dos actos terá de dizer, por exemplo, que roubar, enganar ou torturar é *sempre* correcto se o acto de roubo, engano ou tortura maximizar o bem. Um consequencialista das regras, pelo contrário, reconhecerá princípios que proíbem estes actos. A este respeito, Harsanyi (1977: 58-60) sublinha que a aceitação colectiva de proibições como as indicadas maximiza o bem em virtude de permitir que os agentes formem expectativas definidas quanto à conduta dos outros, incentivando assim a sua cooperação.

Poderá ocorrer-nos que, à semelhança do consequencialismo dos actos, também o consequencialismo das regras tem implicações demasiado contra-intuitivas. Pois, intuitivamente, em algumas circunstâncias possíveis será correcto roubar, enganar ou talvez mesmo torturar. Um código moral que proíba *absolutamente* actos como estes não é coerente com as nossas intuições ponderadas.

A resposta a esta objecção é que, na verdade, as proibições que figuram no código ideal não têm uma força absoluta. Correspondem antes a deveres *prima facie*. Hooker (2000: 134-136) diz-nos que o código ideal inclui uma regra para evitar desastres que suplanta todas as obrigações de não realizar actos de determinados tipos. Mas não será este apelo a uma regra para evitar desastres uma manobra *ad hoc*? Não, pois é bastante fácil compreender que o código moral que teria as melhores consequências há-de incluir uma regra como essa.

Considerem-se dois códigos com proibições de realizar actos como mentir ou quebrar promessas. Um deles, C<sub>1</sub>, inclui também um princípio para evitar desastres que suplanta essas proibições, ao passo que o outro, C<sub>2</sub>, não inclui nenhum princípio desse género. Se este último código conquistasse a aceitação geral, os agentes cederiam a uma ‘veneração das regras’, resultando daí ocorrências desastrosas. Se, pelo contrário, C<sub>1</sub> colhesse tal aceitação, os agentes teriam uma aversão a permitir desastres ainda mais forte do que a sua aversão a mentir ou a quebrar promessas, pelo que estariam dispostos a realizar actos como esses de modo a evitar grandes males. E, como a iminência de desastres é uma situação extraordinária, a existência dessa disposição não criaria graves expectativas negativas na sociedade. É assim muito plausível que o código ideal sancione um princípio para evitar desastres. Na verdade, o consequencialismo das regras explica por que razão é permissível — ou mesmo obrigatório — realizar certos tipos de actos geralmente errados quando fazê-lo é necessário para evitar a ocorrência de desastres. É permissível, por exemplo, quebrar uma promessa trivial de modo a impedir que inocentes sejam torturados, visto que o código ideal aprova uma acção como essa.

Importa reconhecer que a noção de desastre é muito vaga. Afinal, como poderemos determinar se uma má ocorrência decorrente do cumprimento das proibições *prima facie* constituirá um verdadeiro desastre? Podemos dizer, sem dúvida, que o mal resultante da observância

de uma regra, numa dada situação, constituirá um desastre somente se esse mal for muito maior do que o mal resultante da infracção da regra nessa situação. Esta sugestão não atenua significativamente a vagueza, mas talvez o consequencialista das regras não tenha de oferecer uma noção bastante mais precisa de desastre, já que uma noção extremamente precisa seria perniciosa. Um código com estipulações muito detalhas acerca do que constitui um desastre (*i.e.* acerca das condições em que se justifica infringir cada uma das proibições certas de modo a evitar uma má ocorrência) seria muito difícil de interiorizar e de aplicar, pelo que é plausível que o código ideal confie em grande medida ao discernimento do agente a identificação de desastres.

Na esfera da beneficência, o consequencialismo das regras também se revela mais próximo das intuições morais comuns do que o consequencialismo dos actos. À luz desta última teoria, temos a obrigação de fazer sempre tudo o que estiver ao nosso alcance para promover o bem imparcialmente — o que, nas circunstâncias actuais, requer um altruísmo extremo. Como acima se observou, o código ideal é mais leniente. Segundo Hooker (2000: 159-174), o consequencialismo das regras sanciona um princípio de beneficência segundo o qual temos o dever de fazer sacrifícios consideráveis, mas não extremos, em prol dos outros, especialmente dos mais desfavorecidos.<sup>7</sup>

## 9 Além do equilíbrio reflexivo

Que razões, para lá da coerência com as intuições morais ponderadas, teremos para aceitar o consequencialismo das regras? Ou, pelo menos, para o preferir ao consequencialismo dos actos? Nesta última secção, sugerir-se-á que o consequencialismo das regras, além de ter vantagens explicativas e epistémicas assinaláveis sobre o consequencialismo dos actos, ajusta-se a uma visão apelativa da própria moralidade.

<sup>7</sup> Para uma discussão da perspectiva de Hooker, veja-se Galvão (2008: 230-237).



### 9.1 *Propriedades deonticamente relevantes*

Propriedades deonticamente relevantes são aquelas que contribuem para tornar os actos certos ou errados, pelo que é apontando-as que explicaremos o estatuto deontico de actos particulares.<sup>8</sup>

De acordo com o consequencialismo dos actos, há uma só propriedade deonticamente relevante: a de maximizar (imparcialmente) o bem. Qualquer acto correcto é correcto somente em virtude de maximizar o bem. E qualquer acto incorrecto é incorrecto somente em virtude de não ter essa propriedade. Será isto plausível? Consideremos alguns actos incorrectos ou errados, como um certo homicídio de inocentes ou uma certa inflicção escusada de sofrimento a animais. O consequencialista dos actos terá algo de muito desconcertante a dizer: esses actos são errados *simplesmente* porque não geraram o maior bem, vistas as coisas de uma perspectiva imparcial.

Para o consequencialista das regras — ou, pelo menos, para o consequencialista das regras que não procure fundamentar a sua perspectiva numa outra teoria ética —, há uma só propriedade deonticamente relevante fundamental: a de estar de acordo com os princípios optimíficos. No entanto, estes princípios especificam muitas outras propriedades deonticamente relevantes. Confrontado com os actos acima referidos, o consequencialista das regras poderá dizer: esses actos são errados, *em última análise*, porque não estão de acordo com os princípios optimíficos. Contudo, admitindo que os princípios optimíficos proibem matar inocentes e infligir sofrimento escusado a animais, poderá acrescentar: esses actos são errados, *antes de mais*, porque consistem em matar inocentes e em infligir sofrimento escusado a animais. Explicativamente, esta perspectiva afigura-se muito mais apelativa do que aquela que o consequencialista dos actos nos tem para oferecer.

É verdade que os consequencialistas dos actos, cientes de que o seu princípio é muito difícil de aplicar sistematicamente, recomendam a adopção de ‘princípios secundários’, para usar o termo de J. S. Mill (1863: 64-66). Como estes princípios não serão muito diferentes dos princípios optimíficos, haverá uma convergência assinalável, na

<sup>8</sup> A expressão ‘propriedade deonticamente relevante’ é aqui usada como sinónima de ‘right-making property or wrong-making property’.

tomada de decisões, entre aqueles que aceitem o consequencialismo dos actos e aqueles que subscrevam antes o consequencialismo das regras. No entanto, os ‘princípios secundários’ têm um carácter meramente heurístico, pelo que não especificam propriedades deonticamente relevantes para lá da propriedade de maximizar o bem. Assim, esses princípios em nada contribuem para explicar o estatuto deontico dos actos.

### 9.2 *Consequências remotas*

Em muitos casos, não será difícil prever as consequências imediatas das nossas acções. Contudo, as suas consequências remotas, dadas as nossas limitações cognitivas, muitas vezes serão inescrutáveis. Este facto epistémico suscita uma objecção importante ao consequencialismo dos actos: dado que, de um ponto de vista imparcial, as consequências remotas importam tanto como as mais próximas, e sendo as primeiras caracteristicamente inescrutáveis, em demasiados casos importantes será impossível determinarmos que acto terá as melhores consequências. Assim, em demasiados casos importantes, o consequencialismo dos actos deixa os agentes irremediavelmente às escuras quanto ao que fazer.

Uma réplica a esta objecção consistirá em alegar que, de um modo geral, as consequências remotas podem ser ignoradas, visto que estas serão iguais independentemente do que fizermos agora: quer mintamos ou não mintamos agora, por exemplo, tudo será igual dentro de cem anos. James Lenman (Lenman 2000) mostra que esta réplica é profundamente insatisfatória. Pois muitas acções são determinantes de identidades, isto é, determinam quem existirá e quem não existirá nas gerações futuras. Abortos, salvamentos e homicídios são alguns dos exemplos mais óbvios de acções deste tipo. E as acções determinantes de identidades, note-se, têm ramificações causais massivas: os seus efeitos tenderão a aumentar exponencialmente ao longo das gerações. Ou seja, um único aborto, homicídio ou salvamento pode fazer facilmente que, séculos depois, existam milhões de pessoas que de outra forma não teriam existido — e que não existam milhões de pessoas que de outra forma teriam existido. Assim, um único acto determinante de identidades tem tipicamente muitos efeitos significativos que não param de se ramificar ao longo do tempo. Muitos deles serão bons, muitos

outros serão maus — mas, claro, quase todos serão inescrutáveis para o agente. E, como o agente não poderá deixar de ignorar a vastíssima maioria das consequências relevantes de um acto *determinante de identidades*, ficará praticamente às escuras quanto ao que fazer, se aceitar o consequencialismo dos actos.

Será que uma objecção similar ao consequencialismo das regras se revela igualmente forte? Do mesmo modo que as consequências globais de actos particulares são frequentemente inescrutáveis, não serão as consequências globais da aceitação de um código moral, *em virtude de incluírem os efeitos remotos da adesão aos seus princípios*, também inescrutáveis? Se for esse o caso, o consequencialista das regras não estará em melhor posição do que o consequencialista dos actos, pois não conseguirá identificar os princípios optimíficos e, dessa forma, deixar-nos-á sem saber que actos são permissíveis. É de crer, no entanto, que as consequências remotas da aceitação geral de princípios tendem a ser muito mais previsíveis do que as consequências remotas de actos particulares.

Compare-se, por exemplo, as consequências de um acto de matar com as consequências da aceitação geral de um princípio que permita matar. Se alguém mata uma pessoa, poderemos saber com toda a facilidade que isso é imediatamente mau, mas não saberemos se é remotamente mau — talvez esse acto, introduzindo grandes alterações na identidade das pessoas futuras, impeça o nascimento de um novo Estaline, e seja por isso remotamente muito bom. Imagine-se agora que os membros de uma sociedade humana aceitavam um princípio que permitia matar e que, por isso, não tinham a menor aversão a cometer homicídios. Sabemos que isso seria imediatamente muito mau, visto que a estabilidade social ficaria logo comprometida, e sabemos também que, a longo prazo, a inexistência de uma aversão ao homicídio não se tornaria benéfica. Por que razão haveria de ocorrer essa mudança? Do mesmo modo, sabemos que, tanto imediata como remotamente, a aceitação geral de uma regra que proíba matar teria sempre consequências muito boas. Como estamos a supor que a aversão colectiva a matar se manteria estável na sociedade, não será razoável crer que esta, com o passar do tempo, comesse a ter piores resultados.

### 9.3 *A natureza da moralidade*

E se toda a gente fizesse o mesmo? Ou melhor, e se toda a gente se sentisse livre, à vontade, para fazer o mesmo? No pensamento moral comum, esta questão surge recorrentemente quando se pretende avaliar a moralidade de uma acção. Tipicamente, sugere-se ou argumenta-se que, como as consequências de toda a gente se sentir livre para realizar actos de um certo tipo seriam muito más, um certo acto desse tipo é errado. Por exemplo, como as consequências de todos se sentirem livres para fugir aos impostos seriam más, é errado fugirmos aos impostos. O consequencialismo das regras, como observa Hooker (2000: 4-5), vai ao encontro desta ideia comum sobre a natureza da própria moralidade.

Eis outra ideia apelativa sobre a moralidade, que Scanlon (1998) enfatiza: agir moralmente bem é agir de uma forma justificável perante os outros. Ora, podemos acrescentar, quem aja segundo princípios imparcialmente justificáveis estará a agir de uma forma justificável perante os outros. O consequencialismo das regras vai também ao encontro desta ideia, identificando os princípios imparcialmente justificáveis com aqueles que fariam as coisas correr pelo melhor, de um ponto de vista imparcial, se fossem colectivamente interiorizados.

Estas considerações, embora militem a favor do consequencialismo das regras, por si mesmas não justificam a sua aceitação. Outras teorias éticas (e.g. contratualistas ou kantianas) também se ajustam a uma visão apelativa da moralidade. Para determinar se o consequencialismo das regras é superior a essas teorias, há que investigar melhor não só a sua coerência com as intuições morais ponderadas, mas também a sua fertilidade tanto no esclarecimento de questões práticas significativas (e.g. a eutanásia, o estatuto moral dos animais, o castigo estatal) como na resolução dos enigmas teóricos que têm assolado as perspectivas deontológicas e motivado a preferência pelo consequencialismo dos actos.<sup>9</sup>

Pedro Galvão

Faculdade de Letras da Universidade de Lisboa  
LANCOG Group - Centro de Filosofia da Universidade de Lisboa

<sup>9</sup> Agradeço a Ricardo Santos os comentários à versão inicial deste artigo.

## Referências

- Brandt, Richard B. 1979. *A Theory of the Good and the Right*. New York: Prometheus Books.
- Galvão, Pedro. 2008. *Do Ponto de Vista do Universo*. Lisboa: Centro de Filosofia da Universidade de Lisboa.
- Galvão, Pedro. 2018. Agonia e Razões para Agir: Uma Crítica a Parfit. *Philosophica* 52: 31-43.
- Harsanyi, John C. 1953. Cardinal Utility in Welfare Economics and in the Theory of Risk-taking. *The Journal of Political Economy* 61: 434-435.
- Harsanyi, John C. 1955. Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility. *The Journal of Political Economy* 63: 309-321.
- Harsanyi, John C. 1975. Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory. *The American Political Science Review* 69: 594-606.
- Harsanyi, John C. 1977. Morality and the Theory of Rational Behaviour. In *Utilitarianism and Beyond*, ed. by A. Sen and B. Williams. Cambridge: Cambridge University Press, 1982.
- Harsanyi, John C. 1985. Does Reason Tell Us What Moral Code to Follow and, Indeed, to Follow Any Moral Code at All? *Ethics* 96: 42-55.
- Hooker, Brad. 2000. *Ideal Code, Real World*. Oxford: Clarendon Press.
- Hooker, Brad. 2005. Reply to Arneson and McIntyre. *Philosophical Issues* 15: 264-281.
- Hooker, Brad. 2016. Rule Consequentialism. In *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition). Organizada por E. N. Zalta, URL = <<https://plato.stanford.edu/archives/win2016/entries/consequentialism-rule/>>.
- Kant, Immanuel. 1785. *Fundamentação da Metafísica dos Costumes*. Lisboa: Edições 70, 2009.
- Lenman, James. 2000. Consequentialism and Cluelessness. *Philosophy & Public Affairs* 29: 342-370.
- Mill, John Stuart. 1863. *Utilitarismo*. Porto: Porto Editora, 2005.
- Otsuka, Michael. 2009. The Kantian Argument for Consequentialism. In *Essays on Derek Parfit's On What Matters*, ed. by J. Suikkanen and J. Cottingham. Oxford: Wiley-Blackwell.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Clarendon Press.
- Parfit, Derek. 2011a. *On What Matters: Volume One*. Oxford: Oxford University Press.
- Parfit, Derek. 2011b. *On What Matters: Volume Two*. Oxford: Oxford University Press.
- Rawls, John. 1971. *Uma Teoria da Justiça*. Lisboa: Editorial Presença, 1993.
- Rawls, John. 1974. Some Reasons for the Maximin Criterion. *The American Economic Review* 64: 141-146.
- Ridge, Michael. 2006. Introducing Variable-Rate Rule-Utilitarianism. *The Philosophical Quarterly* 56: 242-253.
- Ross, David. 1930. *The Right and the Good*. Oxford: Clarendon Press.
- Ross, Jacob. 2009. Should Kantians Be Consequentialists? In *Essays on Derek Parfit's On What Matters*, ed. by J. Suikkanen and J. Cottingham. Oxford: Wiley-Blackwell.
- Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- Sidgwick, Henry. 1907. *Os Métodos da Ética*. Lisboa: Fundação Calouste Gulbenkian, 2013.

Smart, J. J. C. 1973. *An Outline of a System of Utilitarian Ethics*. In *Utilitarianism: For and Against*. Cambridge: Cambridge University Press.